

УНІВЕРСИТЕТ ІМЕНІ АЛЬФРЕДА НОБЕЛЯ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему

Розробка інтелектуального додатку автоматичного рубрикування тексту

Виконав: здобувач 4 курсу, групи КН-18-1
Спеціальності 122 «Комп'ютерні науки»
Басистий М. О
Керівник: к.т.н., Ковилін Є.Р.

м. Дніпро
2022

АНОТАЦІЯ

Басистий М. О. Розробка інтелектуального додатку автоматичного рубрикування тексту.

Мета: розробка додатку автоматизованого рубрикування тексту.

Предмет: рубрикація тексту на основі словника.

Об'єкт: процеси автоматизованого рубрикування тексту.

Завдання: проведення аналізу існуючих підходів рубрикування тексту, розробити свій підхід рубрикування заснований на аналізі статичних характеристик тексту. І на основі побудованого алгоритму, розробити додаток автоматизованого рубрикування тексту.

Ключові слова: рубрикація, лематизація, комп'ютерна лінгвістика, штучний інтелект, природна мова, полісемія, омонімія, синонімія, текстомайнінг, баєсівський класифікатор.

ЗМІСТ

Вступ.....	6
РОЗДІЛ 1 АНАЛІЗ СТАНУ ПРОБЛЕМИ	8
1.1 Комп’ютерна лінгвістика та її додатки	8
1.2 Етапи і модулі обробки тексту	10
1.3 Огляд текстомайнінгових програм	14
РОЗДІЛ 2 ОПИС ПРЕДМЕТНОЇ ОБЛАСТІ	17
2.1 Сучасні підходи до проектування системи.....	17
2.2 Опис методів реалізації класифікації тексту	18
2.2.1 Побудова алгоритму «навчання з вчителем»	19
2.2.2 Побудова алгоритму байєсівської класифікації.....	19
2.3 Кроки класифікації тексту	21
Розділ 3 ОПИС ПРАКТИЧНОЇ ЧАСТИНИ	23
3.1 Побудова алгоритму додатку	23
3.2 Інтерфейс додатку.....	24
3.3 Обробка даних.....	29
3.4 Аналіз даних	31
ВИСНОВКИ.....	38
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	39
Додаток А.....	41
Додаток Б	48
Додаток В.....	53
Додаток Г	56